Reliability of the 1993 Census Estimates

RELIABILITY OF THE ESTIMATES

An estimate based on a sample survey potentially contains two types of errors— sampling and nonsampling. Sampling errors occur because the estimate is based on a sample, not on the entire universe. Nonsampling errors can be attributed to many sources in the collection and processing of the data. The accuracy of a survey result is affected jointly by the two types of errors. The following is a description of the sampling and nonsampling errors associated with the estimates computed from the 1993 Commodity Flow Survey (CFS).

MEASURES OF SAMPLING VARIABILITY

Because the estimates were based on a sample, exact agreement with the results that would be obtained from a complete census of establishments in the CFS frame using the same enumeration procedure was not expected. How-ever, because each establishment in the Standard Statistical Establishment List (SSEL) in the specified Standard Industrial Classifications (SIC) had a known probability of being selected into the sample, it is possible to estimate the sampling variability of the estimates.

The standard error of the estimate is a measure of the variability among the values of the estimate computed from all possible samples of the same size and design. Thus, it is a measure of the precision with which an estimate from a particular sample approximates the results of a complete enumeration. The coefficient of variation is the standard error of the estimate divided by the value being estimated. It is expressed as a percent. Note that measures of sampling variability, such as the standard error or coefficient of variation, are estimated from the sample and are also subject to sampling variability. Coefficients of variation for number of shipments, dollar value, shipment weight (tons), and ton-miles estimates are shown in tables B- 1 through B- 7 in this appendix. Standard errors for the corresponding percentage estimates are also shown there.

The standard errors and coefficients of variation presented in these tables permit certain confidence statements about the sample estimates. The particular sample used in this survey was one of a large number of samples of the same size that could have been selected using the same design. In about 9 out of 10 (90 percent) of these samples, the estimates would differ from the results of a complete enumeration by less than 1.65 times the standard error of the estimate. In about 19 out of 20 (95 percent) of the samples, the estimates would differ from the result of a complete enumeration by less than twice the standard error of the estimate.

To illustrate the computations involved in the above confidence statements as related to the dollar value estimates, assume that an estimate of shipment value published in table 6 is $10,750 million for a particular commodity and mode of transportation, and that the coefficient of variation for this estimate, as given in appendix A, table B- 6 is 1. 8 percent, or 0. 018. Multiplying $10,750 million by 0.018 yields the standard error, $194 million. Typical practice is to construct a 90- or 95- percent confidence interval. Multiplying $194 million by 1.65 gives $320 million. Therefore, a 90-percent confidence interval is $10,430 million to $11,070 million ($ 10,750 million plus or minus $320 million). If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 9 out of 10 (90 percent) of the intervals would contain the figure obtained from a complete enumeration. Similarly, a 95-percent confidence interval is $10,362 million to $11,138 million ($ 10,750 million plus or minus $388 million).

To illustrate the computations involved related to the percentage estimates, assume that the percentage estimate of shipment value published in table 6 is 25 percent for a particular commodity and mode of transportation, and that the standard error of this estimate, as given in appendix A, table B-6 is 2. 2 percent, or 0. 022. Multiplying 2.2 percent by 1.65 gives 3. 6 percent. So a 90-percent confidence interval is 21.4 percent to 28.6 percent (25 percent plus or minus 3. 6 percent.) If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 9 out of 10 (90 percent) of the intervals would contain the figure obtained from a complete enumeration.

NONSAMPLING ERRORS

As calculated for this report, the standard error and coefficient of variation measures sampling errors but does not measure any systematic biases in the data. Bias is the difference, averaged over all possible samples of the same size and design, between the estimate and the true value being estimated.

In the CFS as in other surveys nonsampling errors can be attributed to many sources: (1) inability to obtain information about all cases in the sample, (2) response errors, (3) definitional difficulties, (4) differences in the interpretation of questions, (5) mistakes in coding or recoding the data obtained, and (6) other errors of collection, response, coverage, and estimation. These nonsampling errors also occur in complete censuses.

Some sources of error are specific to the CFS: (1) Some respondents may have sampled incorrectly when selecting a sample of their documents, (2) some reporters may have used but not reported other units for their measurements— tons instead of pounds, dollars instead of thousands of dollars, etc., (3) on any shipment selected for sample, only the major commodity (by weight) was reported; secondary commodities within shipments were not recorded. Although unlikely, this might lead to a net undercoverage of some secondary commodities. These and other problems could yield a bias of undetermined amount in certain estimates.

Another possible source of bias in estimating the number of shipments, value, shipment weight (tons), and ton-miles is the imputation of missing data and for data which fail edit. Any systematic error in the imputation procedure can introduce bias into the estimates.

Although no direct measurement of the biases due to nonsampling error has been obtained, precautionary steps were taken in all phases of the collection, processing, and tabulation of the data in an effort to minimize their influence.

Biases in the published estimates are due in large part to imputing data for nonrespondents and for data which fail edit. The overall imputation rate for the survey was 30 to 40 percent.